[This Page Intentionally Left Blank]

10.0 ANIMAL WELFARE CONSIDERATIONS (REFINEMENT, REDUCTION, AND REPLACEMENT)

As demonstrated in **Section 6**, *in vitro* NRU basal cytotoxicity test methods cannot be used as replacement assays[1] for rodent acute oral toxicity test methods for hazard classification. However, as described in this section, such test methods can be evaluated for their ability to reduce[2] and refine[3] animal use in the UDP or ATC acute oral toxicity assays. A similar analysis cannot be conducted for the FDP as this test method uses evident toxicity rather than death as the endpoint of interest. The current UDP and ATC test guidelines recommend using information on structurally-related substances and the results of any other toxicity tests (EPA 2002b) to select a starting dose (OECD 2001a; EPA 2002a; OECD 2001d). However, for the purposes of the reduction and refinement evaluation conducted in this section, it was assumed that no information other than 3T3 and NHK NRU test data would be available upon which to base the selection of a starting dose. To determine the extent of animal reduction or refinement that would occur in the UDP and the ATC when using a starting dose based on 3T3 or NHK NRU IC_{50} results rather than the default starting dose, computer models were used to simulate the *in vivo* testing of the reference substances used in the NICEATM/ECVAM validation study.

Section 10.1 lists the regressions that were used with IC_{50} data from the 3T3 and NHK NRU test methods to determine starting doses for the UDP and ATC test methods. **Sections 10.2.1** and **10.3.1** summarize the animal testing procedures described in the current test guidelines for the UDP and the ATC method, respectively. The procedures for using computer software to simulate animal testing of the NICEATM/ECVAM reference substances are then detailed in **Sections 10.2.2** and **10.3.2**. The computer simulations were used to determine the number of animals used and the number of animals that died for each simulated test. The computer simulation modelling was performed using five different dose-mortality (i.e., dose-response)

[1] **Replacement alternative:** A new or modified test method that replaces animals with nonanimal systems or one animal species with a phylogenetically lower one (e.g., a mammal with an invertebrate).
[2] **Reduction alternative:** A new or modified test method that reduces the number of animals required.
[3] **Refinement alternative:** A new or modified test method that refines procedures to lessen or eliminate pain or distress in animals or enhances animal well-being.

slopes since no information on dose-mortality slope was available for the substances tested. To simplify the presentation of results, animal use figures provided in **Sections 10.2.3**, **10.2.4**, **10.3.3**, and **10.3.4** include two of the dose-response slopes. The results for the other three dose-response slopes are provided in **Appendices N** and **Q**. The number of animals used is summarized to show the mean number of animals tested when the default starting dose is used and the mean number of animals used when the NRU-determined starting dose (i.e., from the 3T3 or NHK NRU IC_{50} values used in the indicated regressions) is used. The difference in animal use between the default starting doses and the NRU-based starting doses is referred to as the animal savings. Differences were tested for statistical significance (i.e., $p < 0.05$) using a one-sided Wilcoxon signed ranked test based on the number of substances evaluated. **Sections 10.2** and **10.3** summarize mean animal use by the total number of substances tested and then by the number of substances in each GHS acute oral toxicity category. **Sections 10.2.4** and **10.3.4** provide the mean number of animal deaths compared to the mean number of animals used for each starting dose (i.e., default and NRU-based) to determine whether the NRU-based starting doses result in the refinement of animal use (i.e., reduction in the number of animals that die).

10.1 Use of 3T3 and NHK NRU Test Methods to Predict Starting Doses for Acute Systemic Toxicity Assays

The IC_{50} data from the 3T3 and NHK NRU test methods were used to predict starting doses for acute oral systemic toxicity tests using the following linear regressions of IC_{50}-LD_{50} values presented in **Section 6.2** (see **Table 6-2**):

- the RC millimole regression [Note: The RC millimole regression was developed from the Registry of Cytotoxicity, a database of rat and mouse oral LD_{50} values from RTECS® and IC_{50} values from *in vitro* cytotoxicity assays using multiple cell lines and cytotoxicity endpoints for 347 chemicals with known molecular weights (Halle 1998).]
- the RC rat-only weight regression
- the RC rat-only weight regression excluding substances with specific mechanisms of toxicity other than basal cytotoxicity

Data for the same reference substances were evaluated for each regression and simulated acute systemic toxicity test method. Forty-six substances were evaluated for the 3T3 NRU test method and 47 substances were evaluated for the NHK NRU test method. Of the 72 substances tested, epinephrine bitartrate, colchicine, and propylparaben were excluded because they were removed from the calculation of the RC rat-only weight regression due to the lack of rat oral reference LD_{50} data. The 21 substances with specific mechanisms of toxicity in **Table 6-3** were excluded from all analyses to be consistent with those removed from the RC rat-only weight regression excluding substances with specific mechanisms of toxicity. These substances have known mechanisms of toxicity that are not expected to be active in the 3T3 and NHK cell cultures. Carbon tetrachloride and methanol were excluded from the 3T3 NRU evaluations because no laboratory attained sufficient toxicity in any test for the calculation of an IC_{50}. Carbon tetrachloride was also excluded from the NHK NRU evaluations because no laboratory attained sufficient toxicity in any test for the calculation of an IC_{50}.

10.2 Reduction and Refinement of Animal Use for the UDP

10.2.1 Procedure for *In Vivo* Testing Using the UDP

This section describes the general dosing procedure for the UDP assay (OECD 2001a; EPA 2002a). Although doses, time between doses, and dose progression may be adjusted as necessary, the procedures described reflect the default guidance. Guidance on the type of animals to use, animal housing, clinical observations, etc., are outside the scope of the current discussion and are provided in the test guidelines (see **Appendix M**).

Main Test

The UDP is based on a staircase design in which single animals are dosed in sequence at 48-hour intervals. The outcome of the first animal determines the dose of the next animal. If the first animal dies or is in a moribund state, the dose administered to the next animal is lowered by dividing the original dose by one-half log (i.e., 3.2, which is the default dose progression). If the first animal survives, the dose administered to the next animal is increased by one-half log times the original dose. A dose progression of one-half log unit corresponds to a dose-

mortality (also referred to as "dose-response) slope of 2. The default dose progression can be adjusted if the analyst has prior information upon which to estimate a slope.

The current test guidelines recommend using information on structurally-related substances and the results of any other toxicity tests (EPA 2002b) for the test substance, including *in vitro* cytotoxicity results, to approximate the LD_{50} and the slope of the dose-response curve (OECD 2001a; EPA 2002a). The starting dose is one dose progression step below the analyst's best estimate of the LD_{50}, since the UDP test method has a bias toward the starting dose (i.e., LD_{50} estimate tends to move toward the starting dose). The default starting dose of 175 mg/kg is used if there is no information on which to base a starting dose. The entire default dosing scheme generally uses a dose progression of 3.2, is 1.75, 5.5, 17.5, 55, 175, 550, 1750, and 5000 mg/kg (EPA 2002a) or 1.75, 5.5, 17.5, 55, 175, 550, and 2000 mg/kg (OECD 2001a). Dosing single animals in sequence proceeds until the first of three conditions, referred to as stopping rules, is met:

- three consecutive animals survive at the upper limit (2000 or 5000 mg/kg)
- five reversals occur in any six consecutive animals tested
- four or more animals have followed the first reversal and the specified likelihood-ratios exceed the critical value. For a wide variety of LD_{50} values and dose-mortality slopes, this is satisfied with four to six animals after the first reversal. Three likelihood values are calculated: a likelihood at an LD_{50} point estimate (called the rough estimate or dose-averaging estimate); a likelihood at a value below the point estimate (the point estimate divided by 2.5); and a likelihood at a value above the point estimate (the point estimate multiplied by 2.5). The ratios of the likelihoods are examined to determine whether they exceed a critical value.

If none of these conditions is met, dosing stops after 15 animals have been used.

Limit Test

The UDP test method guidelines include a limit test using three to five animals dosed sequentially at 2000 mg/kg or 5000 mg/kg (OECD 2001a; EPA 2002a). The EPA guideline

for testing at a limit dose of 5000 mg/kg calls for proceeding to the main test if the first animal dosed at 5000 mg/kg dies (EPA 2002a). If the first animal lives, however, two more animals are dosed at 5000 mg/kg. If both animals live, then testing is terminated with LD_{50} > 5000 mg/kg. If one or both animals die, then two more animals are dosed in sequence. As soon as three animals survive, the test is terminated with the conclusion that LD_{50} > 5000 mg/kg. However, as soon as three animals die, the main test is conducted. The OECD guideline for testing at a limit dose of 2000 mg/kg calls for proceeding to the main test if the first animal dosed at 2000 mg/kg dies (OECD 2001a). If the animal lives, however, four more animals are sequentially dosed. Whenever three animals die, the main test is performed. If three or more animals survive, testing is terminated with the conclusion that the LD_{50} > 2000 mg/kg.

10.2.2 Procedure for Computer Simulation Modeling of the UDP

Two thousand simulations of UDP testing were run for each substance, *in vitro* NRU test method, and dose-mortality slope. Because the analysis assumed there was no information upon which to estimate a dose-response slope, the simulation modeling used the default dose progression factor of 3.2. The simulations used 5000 mg/kg as the upper limit dose since this upper limit is commonly used in the United States. If the NRU-based starting dose was 4000 mg/mg or greater, then testing proceeded per the limit test rather than the main test. If, during the dose progression, the next highest dose to be administered was within 4000 mg/kg or greater, then the limit dose of 5000 mg/kg was administered. In the case where a dose one step below the NRU-estimated LD_{50} was used as the starting dose, the other doses administered corresponded to the default doses specified in the test method guidelines (OECD 2001a; EPA 2002a). The simulation modeling procedures also used a lower limit of 1 mg/kg. Thus, if the dose progression fell below 1 mg/kg, then a dose of 1 mg/kg was administered. To estimate animal use by the default method, a starting dose of 175 mg/kg was used; the other doses administered after the default starting dose corresponded to the default doses specified in the test method guidelines (OECD 2001a; EPA 2002a).

The simulation process was performed using SAS® version 8 (SAS 1999) and implements the distributional assumptions underlying the dose-mortality relationship. The lowest dose at

which an animal dies in response to the administration of a toxic substance varies from animal to animal. For an entire population of animals, mortality is assumed to have a log-normal distribution with the mean equal to the log of the true LD_{50}. Sigma (σ), the variability of the simulated population, is the inverse of the slope of the dose-mortality curve. Due to a lack of information for the real dose-mortality curves, the simulations assumed several different values of the slope, but no corresponding changes were made in the dose progression. Dose-mortality slopes of 0.5, 0.8, 2, 4, and 8.3 were chosen since these were used in the simulation modeling studies that evaluated the current version of the UDP guidelines (ICCVAM 2001c).

To model the variability of the NRU IC_{50} values within and between laboratories, the values were log-transformed to normalize the distribution of values for each substance. The mean and variance of these log-transformed values were used to generate a log-normal distribution from which to randomly select an IC_{50} value. The selected NRU IC_{50} value was used with the regressions in two different ways to determine starting doses. One method used the LD_{50} estimated from the IC_{50} and the regression as the starting dose while the other method used the closest default dose lower than the estimated LD_{50} as the starting dose. The results from the latter method are presented in **Section 10.2** since it is the method recommended by the EPA and OECD test guidelines (EPA 2002a; OECD 2001a). Moreover, the UDP is only usable for regulatory purposes if the starting dose is set below the expected LD_{50}. The results obtained when the LD_{50} estimated by the IC_{50} and the regression was used as the starting dose are presented in **Appendix Q.**

The simulation procedure used the following steps for each substance:

1. The LD_{50} value (in mg/kg) from **Table 4-2** was entered as the true LD_{50} value and the choices of assumed slope were entered as the true slope for the dose-mortality curve.
2. An IC_{50} value was selected from a distribution identified by the mean and variance of the IC_{50} values computed from the data to reflect that different laboratories produce different IC_{50} values in different situations (see **Table 5-3** for mean IC_{50} values and standard deviations).

3. The IC_{50} value from Step 2 was used in the regression model being evaluated to compute a predicted LD_{50} value to use as the starting dose.
4. The dosing simulation was run three times: once with the default starting dose of 175 mg/kg, once at the next default dose below the LD_{50} estimated by the regression being evaluated, and once at a dose equal to that of the LD_{50} estimated by the regression being evaluated.
5. For each simulated trial (each substance and starting dose), the dosing simulation works similarly. In each trial, the animals are dosed sequentially; therefore for each animal(i) there is a corresponding dose(i) that is administered to the animal. For the first animal in each trial, it is the starting dose for that trial. For each subsequent animal, the dose is dependent on the previous dose and the previous animal's response as described in **Section 10.2.1**. For animal(i), the probability of response is computed with the cumulative log-normal distribution at the dose administered. That is, $P(response) = P(x < \log[dose(i)])$ where $x \sim N(\mu,\sigma)$ and μ is the log of the true LD_{50} value and σ is the inverse of the assumed slope of the dose-mortality curve. This probability is used to sample one observation from a binomial distribution with this probability of success.
6. Dosing simulation is stopped once one of the stopping rules is satisfied.

Steps 2-6 were repeated 2000 times in order to compute an average animal use for each method evaluated.

10.2.3 Animal Savings for the UDP When Using 3T3 and NHK NRU-Based Starting Doses

10.2.3.1 *The Effect of Dose-Response Slope on Animal Use*

As described in **Section 10.2.2**, the simulation modeling of animal use for the UDP assumed five different dose-mortality slopes to assess animal use under various conditions of population variability. **Table 10-1** shows that the number of animals used for the UDP decreases with increasing slope for both the default starting dose and the NRU-determined starting dose based on the RC millimole regression. The NRU-determined starting dose was

the next default dose lower than the regression-estimated LD_{50}. For example, since the LD_{50} predicted for cadmium chloride by the 3T3 NRU IC_{50} with the RC millimole regression was 16 mg/kg, the starting dose was 1.75 mg/kg (i.e., the next default dose below the predicted LD_{50}). This approach is consistent with the UDP test method guidelines (OECD 2001a; EPA 2002a) as a means for reducing the number of animals that might experience pain and suffering from treatment (i.e., as a test method refinement). The approach also overcomes the nonconservative bias of the UDP, which tends to yield an LD_{50} close to the starting dose.

Table 10-1 Change in Animal Use[1] with Dose-Response Slope for the UDP[2]

Dose-Response Slope	With Default Starting Dose[1,3]	With NRU-Based Starting Dose[1,4]	Animals Saved[5]
3T3 NRU Test Method			
0.5	10.30 ± 0.13	9.43 ± 0.15	0.88* (8.5%)
0.8	10.34 ± 0.17	9.36 ± 0.18	0.98* (9.4%)
2.0	9.77 ± 0.21	8.79 ± 0.22	0.97* (10.0%)
4.0	8.96 ± 0.25	8.03 ± 0.27	0.93* (10.4%)
8.3	8.11 ± 0.26	7.20 ± 0.30	0.91* (11.2%)
NHK NRU Test Method			
0.5	10.31 ± 0.12	9.57 ± 0.17	0.74* (7.1%)
0.8	10.38 ± 0.16	9.47 ± 0.19	0.91* (8.8%)
2.0	9.75 ± 0.20	8.93 ± 0.23	0.82* (8.4%)
4.0	8.94 ± 0.24	8.14 ± 0.28	0.80* (9.0%)
8.3	8.12 ± 0.25	7.33 ± 0.30	0.79* (9.7%)

[1]Numbers are mean numbers of animals with standard errors for 2000 simulations for 46 substances for the 3T3 NRU test method and 47 substances for the NHK NRU test method. Although the simulations used whole animals, averaging the results produced fractional numbers of animals. The slight differences in the number of animals used for the default starting dose at the same dose-response slope reflect different simulation runs. Limit dose = 5000 mg/kg.
[2]OECD (2001a); EPA (2002a).
[3]Default starting dose = 175 mg/kg.
[4]Starting dose = next lower default dose to NRU-predicted LD_{50}, which was calculated using the geometric mean of the laboratory geometric mean NRU IC_{50} values in the RC millimole regression: $\log LD_{50}$ (mmol/kg) = 0.435 $\log IC_{50}$ (mM) + 0.625.
[5]Difference between mean animal use with default starting dose and mean animal use with NRU-based starting dose. All differences denoted by * were statistically significant (i.e., $p < 0.05$) by a one-sided Wilcoxon signed rank test. Percentage difference is shown in parentheses.

Table 10-1 shows that, for each dose-response slope, the mean number of animals saved was statistically significant (i.e., $p < 0.05$) when compared to mean animal use for the default

starting dose. When expressed as a percentage of the default animal use, animal savings also generally increased with increasing slope.

To simplify the presentation of animal savings and comparison of the various regressions and starting doses, the results of subsequent analyses presented in **Section 10.2.3** will be limited to slopes of 2 and 8.3. The slope of 2 is the default slope used for the calculation of LD_{50} by the UDP method (OECD 2001a; EPA 2002a). Animal savings results for the other dose-mortality slopes are presented in **Appendices N1-N3.** Although using the next lower default dose to the NRU-determined LD_{50} value overcomes the bias of the UDP toward the starting dose (OECD 2001a, EPA 2002a) and is the appropriate approach for regulatory use, animal savings results using the estimated LD_{50} as the starting dose were also calculated (see **Appendix Q**).

10.2.3.2 *Mean Animal Use from UDP Simulations for Testing the NICEATM/ECVAM Reference Substances – Comparison of Regressions and 3T3 and NHK NRU Test Methods*

Table 10-2 shows the mean animal use for simulated UDP of the testing the set of NICEATM/ECVAM reference substances described in **Section 10.1**. Mean animal use is shown for default starting dose and for starting doses that were one default dose lower than the LD_{50} predicted from the *in vitro* NRU test methods and the regressions (shown in **Table 6-2**) evaluated in **Section 6.3** for prediction of GHS acute oral toxicity category. The difference in animal use between the two starting doses is the mean animal savings produced by using the starting dose based on the *in vitro* NRU test methods. All differences (i.e., mean animal savings) were statistically significant (i.e., $p < 0.05$) by a one-sided Wilcoxon signed rank test. Mean animal savings ranged from 0.79 to 1.16 (8.4 to 12.7%) animals depending upon the NRU test method, regression, and dose-response slope. The lowest mean animal savings were obtained for the RC millimole regression (0.82 [8.4%] to 0.97 [10.0%] animals for the various test methods and dose-response slopes) and the highest mean animal savings were obtained with the RC rat-only regression excluding substances with specific mechanisms of toxicity other than basal cytotoxicity (1.00 [12.2%] to 1.16 [11.8%] animals).

Table 10-2 Mean Animal Use[1] for the UDP[2] Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with Various Regressions

Assay/Regression	With Default Starting Dose[3]	With NRU-Based Starting Dose[4]	Animals Saved[5]	With Default Starting Dose[3]	With NRU-Based Starting Dose[5]	Animals Saved[5]	Accuracy[6]
3T3 NRU Test Method	**Dose-Response Slope = 2**			**Dose-Response Slope = 8.3**			
RC millimole[6]	9.77 ± 0.21	8.79 ± 0.22	0.97* (10.0%)	8.11 ± 0.26	7.20 ± 0.30	0.91* (11.2%)	26%
RC rat-only weight[7]	9.79 ± 0.21	8.66 ± 0.22	1.13* (11.6%)	8.14 ± 0.25	7.11 ± 0.29	1.03* (12.7%)	35%
RC rat-only weight excluding substances with specific mechanisms of toxicity[8]	9.80 ± 0.20	8.64 ± 0.23	1.16* (11.8%)	8.16 ± 0.25	7.08 ± 0.31	1.08* (13.3%)	46%
NHK NRU Test Method	**Dose-Response Slope = 2**			**Dose-Response Slope = 8.3**			
RC millimole[6]	9.75 ± 0.20	8.93 ± 0.23	0.82* (8.4%)	8.12 ± 0.25	7.33 ± 0.30	0.79* (9.7%)	28%
RC rat-only weight[7]	9.77± 0.20	8.83 ± 0.23	0.94* (9.6%)	8.13 ± 0.25	7.25 ± 0.30	0.88* (10.9%)	30%
RC rat-only weight excluding substances with specific mechanisms of toxicity[8]	9.78 ± 0.20	8.73 ± 0.24	1.05* (10.7%)	8.15 ± 0.25	7.15 ± 0.32	1.00* (12.2%)	38%

[1]Numbers are mean numbers of animals and standard errors for 2000 simulations for each of 46 substances for the 3T3 NRU test method and 47 substances for the NHK NRU test method. Although the simulations used whole animals, averaging the results produced fractional numbers of animals. The slight differences in the number of animals used for the default starting dose at the same dose-response slope reflect different simulation runs.

[2]OECD (2001a); EPA (2002a).

[3]Default starting dose = 175 mg/kg.

[4]Starting dose = one default dose lower than the NRU-predicted LD_{50} calculated using the geometric mean of the laboratory geometric mean NRU IC_{50} values in the specified regression.

[5]Difference between mean animal use with default starting dose and mean animal use with NRU-based LD_{50}. Differences denoted by * were statistically significant (i.e., $p < 0.05$) by a one-sided Wilcoxon signed rank test. Percentage difference is shown in parentheses.

[6]Proportion of substances for which the GHS acute oral toxicity category (UN 2005) predicted by the *in vitro* NRU test methods matched the *in vivo* category (from **Tables 6-4** to **6-6**).

[7]$\log LD_{50}$ (mmol/kg) = 0.435 $\log IC_{50}$ (mM) + 0.625.

[8]$\log LD_{50}$ (mg/kg) = 0.372 $\log IC_{50}$ (µg/mL) + 2.024.

[9]$\log LD_{50}$ (mg/kg) = 0.357 $\log IC_{50}$ (µg/mL) + 2.194.

Table 10-2 also shows that animal savings increased with the accuracy of the GHS acute oral toxicity category predictions (see **Section 6.3**).

10.2.3.3 *Animal Savings for the UDP by Toxicity Category Using 3T3 and NHK NRU-Based Starting Doses*

Tables 10-3 through **10-5** show mean animal use and mean animal savings for the UDP for the default starting dose and the NRU-determined starting dose with the test substances grouped by GHS acute oral toxicity category (UN 2005). The data come from the same analyses as the data provided in **Table 10-2**. NRU-determined starting doses were based on the:

- RC millimole regression (**Table 10-3**).
- RC rat-only weight regression (**Table 10-4)**
- RC rat-only weight regression excluding substances with specific mechanisms of toxicity other than basal cytotoxicity (**Table 10-5)**

Consistencies noted in the mean animal savings data provided in the tables included:

- For each *in vitro* NRU cytotoxicity test method and regression, animal savings were statistically significant for substances in the $2000 < LD_{50} \leq 5000$ mg/kg and $LD_{50} > 5000$ mg/kg toxicity categories.
- For substances with $LD_{50} \leq 5$ mg/kg, the NHK NRU test method with each regression used slightly more animals than the default method (i.e., mean animal savings were negative). The 3T3 NRU test method produced nonsignificant animal savings of 0.31 (2.9%) to 0.95 (8.1%) animal for these substances.

For substances with $50 < LD_{50} \leq 300$ mg/kg, all test methods and regressions produced little to no animal savings.

Animal Savings for the UDP by Toxicity Category Using 3T3 and NHK NRU-Based Starting Doses with the RC Millimole Regression

Table 10-3 shows the animal savings by GHS toxicity category for the *in vitro* NRU cytotoxicity test methods used with the RC millimole regression. Mean animal savings were

statistically significant (i.e., $p < 0.05$) by a one-tailed Wilcoxon signed rank test for the following GHS toxicity categories, test methods, and dose-response slopes:

- $5 < LD_{50} \leq 50$ mg/kg for the NHK NRU at dose-response slope = 2 (0.86 [9.2%] animals)
- $2000 < LD_{50} \leq 5000$ mg/kg for both NRU test methods and both dose-response slopes (1.25 [13.7%] to 1.52 [14.1%] animals)
- $LD_{50} > 5000$ mg/kg for both NRU test methods and both dose-response slopes (1.35 [14.2%] to 1.70 [25.4%] animals)

For the 3T3 NRU and NHK NRU test methods, mean animal savings were similar for most toxicity categories at both dose-response slopes, with the mean savings for the 3T3 NRU slightly higher than that for the NHK NRU. For the dose-response slope of 2, mean animal savings for the 3T3 NRU test method (for the various toxicity categories) ranged from -0.09 (-1.0%) to 1.54 (16.1%) animals while mean animal savings for the NHK NRU test method ranged from -0.25 (-2.2%) to 1.45 (13.5%) animals. For the dose-response slope of 8.3, animal savings for the 3T3 NRU test method ranged from 0.004 (0.05%) to 1.70 (25.4%) animals while mean animal savings for the NHK NRU test method ranged from -0.11 (-1.5%) to 1.45 (21.8%) animals.

For both *in vitro* NRU cytotoxicity test methods, no mean animal savings (≤ 0.09 animal) were observed for substances with $50 < LD_{50} \leq 300$ mg/kg. This category includes the default starting dose of 175 mg/kg. Animal savings were not expected for this category since savings were determined by comparing animal use with the NRU-based starting dose with animal use at the default starting dose. For the 3T3 NRU, no animal savings (-0.9 to 0.004 animals) were also observed for substances with $5 < LD_{50} \leq 50$ mg/kg. For the NHK NRU test method, animal use actually increased slightly compared to the default starting dose (-0.25 to -0.09 animals) for substances with $LD_{50} \leq 5$ mg/kg. Animal savings for relatively high toxicity substances were noted for those in the $LD_{50} \leq 5$ mg/kg category for the 3T3 NRU (0.78 [7.3%] to 0.95 [8.1%] animals) and in the $5 < LD_{50} \leq 50$ mg/kg category for the NHK NRU (0.86 [9.2%] to 0.87 [10.5%] animals). Only the 0.86 (9.2%) animal savings for the dose-response slope of 2 (NHK NRU) were statistically significant.

Table 10-3 Animal Use[1] for the UDP[2] by GHS Toxicity Category[3] Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Millimole Regression[4]

Toxicity Category[3]	Number of Reference Substances	Dose-Response Slope = 2			Dose-Response Slope = 8.3			Accuracy[8]
		With Default Starting Dose[5]	With NRU-Based Starting Dose[6]	Animals Saved[7]	With Default Starting Dose[5]	With NRU-Based Starting Dose[6]	Animals Saved[7]	
		3T3 NRU Test Method						
$LD_{50} \leq 5$ mg/kg	7	11.76 ± 0.16	10.8 ± 0.64	0.95 (8.1%)	10.65 ± 0.48	9.87 ±0.74	0.78 (7.3%)	0%
$5 < LD_{50} \leq 50$ mg/kg	6	9.06 ± 0.18	9.15 ± 0.72	-0.09 (-1.0%)	8.04 ± 0.24	8.04 ± 0.78	0.004 (0.05%)	17%
$50 < LD_{50} \leq 300$ mg/kg	6	7.70 ± 0.23	7.61 ± 0.18	0.09 (1.2%)	6.63 ± 0.35	6.59 ± 0.26	0.03 (0.5%)	67%
$300 < LD_{50} \leq 2000$ mg/kg	6	8.76 ± 0.34	7.91 ± 0.06	0.84 (9.6%)	7.30 ± 0.35	6.69 ± 0.20	0.61 (8.3%)	100%
$2000 < LD_{50} \leq 5000$ mg/kg	11	10.75 ± 0.08	9.23 ± 0.20	1.52* (14.1%)	9.16 ± 0.26	7.81 ± 0.34	1.36* (14.8%)	0%
$LD_{50} > 5000$ mg/kg	10	9.59 ± 0.27	8.05 ± 0.39	1.54* (16.1%)	6.69 ± 0.37	4.99 ± 0.45	1.70* (25.4%)	10%
		NHK NRU Test Method						
$LD_{50} \leq 5$ mg/kg	7	11.54 ± 0.25	11.79 ± 0.50	-0.25 (-2.2%)	10.63 ± 0.49	10.72 ± 0.54	-0.09 (-0.8%)	0
$5 < LD_{50} \leq 50$ mg/kg	6	9.34 ± 0.24	8.48 ± 0. 24	0.86* (9.2%)	8.22 ± 0.31	7.35 ± 0.36	0.87 (10.5%)	50%
$50 < LD_{50} \leq 300$ mg/kg	6	7.82 ± 0.22	7.88 ± 0.26	-0.06 (-0.7%)	6.92 ± 0.38	7.02 ± 0.43	-0.11 (-1.5%)	50%
$300 < LD_{50} \leq 2000$ mg/kg	6	8.74 ± 0.34	7.93 ± 0.06	0.81 (9.3%)	7.31 ± 0.34	6.71 ± 0.23	0.60 (8.2%)	100%
$2000 < LD_{50} \leq 5000$ mg/kg	11	10.73 ± 0.08	9.29 ± 0.20	1.45* (13.5%)	9.13 ± 0.25	7.88 ± 0.33	1.25* (13.7%)	9%
$LD_{50} > 5000$ mg/kg	11	9.52 ± 0.28	8.17 ± 0.41	1.35* (14.2%)	6.64 ± 0.35	5.19 ± 0.44	1.45* (21.8%)	0%

[1]Numbers are mean numbers of animals used and standard errors for 2000 simulations for each substance with a limit dose of 5000 mg/kg. Although the simulations used whole animals, averaging the results produced fractional numbers of animals. Results are provided for 46 substances in the 3T3 NRU test method and 47 substances in the NHK NRU test method categorized using the initial LD_{50} values from **Table 3-2**. The slight differences in the number of animals used for the default starting dose at the same dose-response slope reflect different simulation runs.

[2]OECD (2001a); EPA (2002a).

[3]GHS-Globally Harmonized System of Classification and Labelling of Chemicals with LD_{50} in mg/kg (UN 2005).

[4]RC millimole regression is $\log LD_{50}$ (mmol/kg) = 0.435 $\log IC_{50}$ (mM) + 0.625.

[5]Default starting dose = 175 mg/kg.

[6]Starting dose was one default dose lower than the predicted LD_{50} calculated using the geometric mean of the laboratory geometric mean NRU IC_{50} values in the RC millimole regression.

[7]Difference between mean animal use with default starting dose and mean animal use with NRU predicted LD_{50}. Differences marked by * are statistically significant ($p < 0.05$) by a one-sided Wilcoxon signed rank test. Percentage difference shown in parentheses

[8]Proportion of substances for which the GHS acute oral toxicity category (UN 2005) predicted by the *in vitro* NRU test methods matched the *in vivo* category (from **Table 6-4**).

Table 10-3 also shows that mean animal savings did not correlate with the accuracy of the GHS acute oral toxicity category predictions. Substances in categories with the lowest accuracy produced the highest animal savings. Accuracy was the lowest (0 - 10%) for GHS acute oral toxicity category prediction for substances with $LD_{50} > 5000$ mg/kg, but animal savings (1.35 - 1.70) were the highest. Animal savings (0.60 - 0.84 animals) for substances with $300 \leq LD_{50} \leq 2000$ mg/kg, which had 100% accuracy for GHS acute oral toxicity category prediction, were similar to animal savings (0.78 - 0.95 animals) for substances in the $LD_{50} < 5$ mg/kg category (for the 3T3 NRU), which had 0% accuracy. Perhaps the difference between the predicted starting dose and the true LD_{50} vs. the difference between the default starting dose and the true LD_{50} has more influence on animal savings that the accuracy of the LD_{50} prediction.

Animal Savings for the UDP by Toxicity Category Using 3T3 and NHK NRU-Based Starting Doses with the RC Rat-Only Weight Regression

Table 10-4 shows the mean animal savings by GHS toxicity category for the *in vitro* NRU cytotoxicity test methods used with the RC rat-only weight regression. A comparison of mean animal savings, category for category, with the RC millimole regression, indicates that, in most cases, animal savings were slightly higher for the RC rat-only weight regression. For the RC rat-only weight regression, the mean differences between animal use for the default starting dose and mean animal use with the NRU-determined starting dose were statistically significant (i.e., $p < 0.05$) by a one-sided Wilcoxon signed rank test for the following GHS toxicity categories, test methods, and dose-response slopes:

- $300 < LD_{50} \leq 2000$ mg/kg for the NHK NRU at dose-response slope = 2 (0.86 [9.8%] animals)
- $2000 < LD_{50} \leq 5000$ mg/kg for both NRU test methods and both dose-response slopes (1.50 [16.4%] to 1.91 [17.7%] animals)
- $LD_{50} > 5000$ mg/kg for both NRU test methods and both dose-response slopes (1.45 [15.2%] to 1.73 [25.9%] animals)

Table 10-4 Animal Use[1] for the UDP[2] by GHS Toxicity Category[3] Using Starting Doses Based on the NRU Test Methods with the RC Rat-Only Weight Regression[4]

		Dose-Response Slope = 2			Dose-Response Slope = 8.3			
Toxicity Category[3]	**Number of Reference Substances**	**With Default Starting Dose[5]**	**With NRU-Based Starting Dose**	**Animals Saved[7]**	**With Default Starting Dose[5]**	**With NRU-Based Starting Dose**	**Animals Saved[7]**	**Accuracy[8]**
		3T3 NRU Test Method						
$LD_{50} \leq 5$ mg/kg	4	11.75 ± 0. 16	10.85 ± 0.61	0.89 (7.6%)	10.66 ± 0.48	9.93 ± 0.71	0.73 (6.8%)	0%
$> 5 < LD_{50} \leq 50$ mg/kg	7	9.14 ± 0.17	8.80 ± 0.54	0.34 (3.7%)	8.12 ± 0.27	7.76 ± 0.59	0.36 (4.5%)	17%
$> 50 < LD_{50} \leq 300$ mg/kg	5	7.75 ± 0.22	7.60 ± 0.10	0.15 (1.9%)	6.71 ± 0.32	6.66 ± 0.23	0.05 (0.8%)	67%
$> 300 < LD_{50} \leq 2000$ mg/kg	9	8.75 ± 0.33	7.89 ± 0.07	0.86* (9.8%)	7.29 ± 0.35	6.68 ± 0.21	0.61 (8.4%)	100%
$> 2000 < LD_{50} \leq 5000$ mg/kg	9	10.81 ± 0.08	8.90 ± 0.28	1.91* (17.7%)	9.18 ± 0.26	7.48 ± 0.42	1.70* (18.5%)	0%
> 5000 mg/kg	12	9.59 ± 0.27	7.96 ± 0.40	1.63* (17.0%)	6.69 ± 0.37	4.96 ± 0.45	1.73* (25.9%)	10%
		NHK NRU Test Method						
$LD_{50} \leq 5$ mg/kg	4	11.58 ± 0.23	11.66 ± 0.44	-0.08 (-0.7%)	10.66 ± 0.48	10.59 ± 0.53	0.07 (0.6%)	0
$> 5 < LD_{50} \leq 50$ mg/kg	7	9.33 ± 0.26	8.39 ± 0.27	0.94 (10.1%)	8.20 ± 0.31	7.36 ± 0.38	0.84 (10.3%)	50%
$> 50 < LD_{50} \leq 300$ mg/kg	5	7.84 ± 0.21	7.93 ± 0.25	-0.09 (-1.1%)	6.94 ± 0.37	7.09 ± 0.41	-0.15 (-2.2%)	50%
$> 300 < LD_{50} \leq 2000$ mg/kg	9	8.74 ± 0.34	7.92 ± 0.06	0.82 (9.3%)	7.31 ± 0.34	6.71 ± 0.23	0.60 (8.2%)	100%
$> 2000 < LD_{50} \leq 5000$ mg/kg	9	10.77 ± 0.07	9.07 ± 0.24	1.70*(15.8%)	9.14 ± 0.25	7.64 ± 0.37	1.50* (16.4%)	9%
$LD_{50} > 5000$ mg/kg	13	9.52 ± 0.28	8.07 ± 0.40	1.45*(15.2%)	6.64 ± 0.35	5.09 ± 0.42	1.55* (23.3%)	0%

[1]Numbers are mean number of animals used and standard errors for 2000 simulations for each substance with a limit dose of 5000 mg/kg. Although the simulations used whole animals, averaging the results produced fractional numbers of animals. Results are provided for 46 substances in the 3T3 NRU test method and 47 substances in the NHK NRU test method categorized using the reference LD_{50} values from **Table 4-2**. The slight differences in the number of animals used for the default starting dose at the same dose-response slope reflect different simulation runs.

[2]OECD (2001a); EPA (2002a).

[3]GHS-Globally Harmonized System of Classification and Labelling of Chemicals with LD_{50} in mg/kg (UN 2005).

[4]From **Table 6-2**; log LD_{50} (mg/kg) = 0.372 log IC_{50} (µg/mL) + 2.024

[5]Default starting dose = 175 mg/kg.

[6]Starting dose was one default dose lower than NRU-predicted LD_{50} calculated using the geometric mean of the laboratory geometric mean NRU IC_{50} values in the RC rat-only regression.

[7]Difference between mean animal use with default starting dose and mean animal use with NRU predicted LD_{50}. Differences marked by * were statistically significant (i.e., $p < 0.05$) by a one-sided Wilcoxon signed rank test. Percent difference is shown in parentheses.

[8]Proportion of substances for which the GHS acute oral toxicity category (UN 2005) predicted by the *in vitro* NRU test methods matched the *in vivo* category (from **Table 6-5**).

For the dose-response slope of 2, mean animal savings (for the various toxicity categories) for the 3T3 NRU test method ranged from 0.15 (1.9%) to 1.91 (17.7%) animals while mean animal savings for the NHK NRU test method ranged from -0.09 (-1.1%) to 1.70 (15.8%) animals. For the dose-response slope of 8.3, animal savings for the 3T3 NRU test method ranged from 0.05 (0.8%) to 1.73 (25.9%) animals while animal savings for the NHK NRU test method ranged from -0.15 (-2.2%) to 1.55 (23.3%) animals.

For both *in vitro* NRU cytotoxicity test methods, no mean animal savings (≤ 0.15 animal) were observed for substances with $50 < LD_{50} \leq 300$ mg/kg. This category includes the default starting dose of 175 mg/kg. Animal savings were not expected for this category since savings were determined by comparing animal use with the NRU-based starting dose with animal use at the default starting dose. For the NHK NRU, no animal savings (-0.08 to 0.07 animals) were also observed for substances with $LD_{50} \leq 5$ mg/kg. Animal savings for relatively high toxicity substances were noted in the $LD_{50} \leq 5$ mg/kg category for the 3T3 NRU (0.73 [6.8%] to 0.89 [7.6%] animals) and in the $5 < LD_{50} \leq 50$ mg/kg category for the NHK NRU (0.84 [10.3%] to 0.94 [10.1%] animals), but these savings were not statistically significant.

Table 10-4 also shows that mean animal savings did not correlate with the accuracy of the GHS acute oral toxicity category predictions (see **Section 6.3**). The toxicity categories with the highest animal savings had low accuracy. Substances in the $2000 < LD_{50} \leq 5000$ mg/kg and $LD_{50} > 5000$ mg/kg categories had very low accuracy (0 - 10%) for GHS acute oral toxicity category prediction, but the animal savings were higher than for the other categories (1.45-1.91). Additionally, animal savings (0.61 - 0.86 animals) for substances with $300 \leq LD_{50} \leq 2000$ mg/kg, which had 100% accuracy for GHS acute oral toxicity category prediction, were similar to animal savings (0.73 - 0.89 animals) for substances in the $LD_{50} <$ 5 mg/kg category (for the 3T3 NRU), which had 0% accuracy. Perhaps the difference between the predicted starting dose and the true LD_{50} vs. the difference between the default starting dose and the true LD_{50} has more influence on animal savings than the accuracy of the LD_{50} prediction.

Animal Savings for the UDP by Toxicity Category Using 3T3 and NHK NRU-Based Starting Doses with the RC Rat-Only Weight Regression Excluding Substances with Specific Mechanisms of Action

Table 10-5 shows the mean animal savings by GHS toxicity category for the *in vitro* NRU cytotoxicity test methods used with the RC rat-only weight regression excluding substances with specific mechanisms of toxicity other than basal cytotoxicity. For substances in the categories for LD_{50} > 2000 mg/kg, mean animal savings for the RC rat-only weight regression excluding substances with specific mechanisms of toxicity other than basal cytotoxicity were slightly higher than those for the RC rat-only weight regression and those for the RC millimole regression. Mean differences between animal use for the default starting dose and mean animal use with the NRU-determined starting dose were statistically significant (i.e., $p < 0.05$) by a one-sided Wilcoxon signed rank test for the following GHS toxicity categories, test methods, and dose-response slopes:

- $5 < LD_{50} \leq 50$ mg/kg for the NHK NRU at dose-response slope = 2 (0.98 [10.6%] animals)
- $300 < LD_{50} \leq 2000$ mg/kg for both NRU test methods and at dose-response = 2 (1.00 [11.4%] animals for the 3T3 NRU and 0.90 [10.3%] animals for the NHK NRU)
- $2000 < LD_{50} \leq 5000$ mg/kg for both NRU test methods and both dose-response slopes (1.75 [19.1%] to 2.22 [20.5%] animals)
- LD_{50} > 5000 mg/kg for both NRU test methods and both dose-response slopes (1.77 [18.6%] to 2.01 [30.1%] animals)

Mean animal savings for the 3T3 NRU and NHK NRU test methods were similar for each toxicity category and dose-response slope, with the 3T3 NRU test method producing slightly higher mean animal savings in most cases. For the dose-response slope of 2, mean animal savings across the various toxicity categories for the 3T3 NRU ranged from -0.02 (-0.2%) to 2.22 (20.5%) animals while mean animal savings for the NHK NRU ranged from -0.35 (-3.0%) to 1.98 (18.3%) animals.

Table 10-5 Animal Use[1] for the UDP[2] By GHS Toxicity Category[3] Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Rat-Only Weight Regression Excluding Substances with Specific Mechanisms of Toxicity[4]

Toxicity Category[3]	Number of Reference Substances	Dose-Response Slope = 2			Dose-Response Slope = 8.3			Accuracy[8]
		With Default Starting Dose[5]	With NRU-Based Starting Dose	Animals Saved[7]	With Default Starting Dose[5]	With NRU-Based Starting Dose	Animals Saved[7]	
		3T3 NRU Test Method						
$LD_{50} \leq 5$ mg/kg	4	11.68 ± 0.17	11.26 ± 0.55	0.42 (3.6%)	10.62 ± 0.48	10.31 ± 0.67	0.31 (2.9%)	0%
$> 5 < LD_{50} \leq 50$ mg/kg	7	9.05 ± 0.13	9.03 ± 0.55	0.02 (0.3%)	8.07 ± 0.25	7.92 ± 0.59	0.15 (1.9%)	14%
$> 50 < LD_{50} \leq 300$ mg/kg	5	7.82 ± 0.18	7.84 ± 0.15	-0.02 (-0.2%)	6.93 ± 0.31	6.99 ± 0.29	-0.06 (-0.9%)	80%
$> 300 < LD_{50} \leq 2000$ mg/kg	9	8.81 ± 0.35	7.81 ± 0.06	1.00* (11.4%)	7.31 ± 0.37	6.58 ± 0.18	0.73 (10.0%)	78%
$> 2000 < LD_{50} \leq 5000$ mg/kg	9	10.84 ± 0.07	8.62 ± 0.23	2.22* (20.5%)	9.18 ± 0.26	7.19 ± 0.37	2.00* (21.8%)	67%
> 5000 mg/kg	12	9.59 ± 0.27	7.71 ± 0.40	1.88* (19.6)%	6.69 ± 0.37	4.68 ± 0.46	2.01* (30.1%)	25%
		NHK NRU Test Method						
$LD_{50} \leq 5$ mg/kg	4	11.55 ± 0.23	11.90 ± 0.32	-0.35(-3.0%)	10.66 ± 0.48	10.83 ± 0.45	-0.18 (-1.6%)	0
$> 5 < LD_{50} \leq 50$ mg/kg	7	9.28 ± 0.25	8.30 ± 0.28	0.98* (10.6%)	8.19 ± 0.32	7.30 ± 0.36	0.89 (10.9%)	14%
$> 50 < LD_{50} \leq 300$ mg/kg	5	7.87 ± 0.20	8.03 ± 0.24	-0.16 (-2.0%)	7.08 ± 0.34	7.26 ± 0.40	-0.19 (-2.6%)	60%
$> 300 < LD_{50} \leq 2000$ mg/kg	9	8.76 ± 0.33	7.86 ± 0.06	0.90* (10.3%)	7.31 ± 0.34	6.61 ± 0.22	0.69 (9.5%)	89%
$> 2000 < LD_{50} \leq 5000$ mg/kg	9	10.82 ± 0.07	8.84 ± 0.26	1.98* (18.3%)	9.15 ± 0.25	7.41 ± 0.39	1.75* (19.1%)	44%
$LD_{50} > 5000$ mg/kg	13	9.52 ± 0.28	7.75 ± 0.43	1.77* (18.6%)	6.64 ± 0.35	4.76 ± 0.44	1.88* (28.4%)	15%

[1]Numbers are mean number of animals used and standard errors for 2000 simulations for each substance with a limit dose of 5000 mg/kg. Although the simulations used whole animals, averaging the results produced fractional numbers of animals. Results are provided for 46 substances in the 3T3 NRU test method and 47 substances in the NHK NRU test method categorized using the reference LD_{50} values from **Table 4-2**. The slight differences in the number of animals used for the default starting dose at the same dose-response slope reflect different simulation runs.

[2]OECD (2001a); EPA (2002a).

[3]GHS-Globally Harmonized System of Classification and Labelling of Chemicals with LD_{50} in mg/kg (UN 2005).

[4]From **Table 6-2**; log LD_{50} (mg/kg) = 0.357 log IC_{50} (μg/mL) + 2.194.

[5]Default starting dose = 175 mg/kg.

[6]Starting dose = One default dose lower than NRU-predicted LD_{50} calculated using the geometric mean of laboratory mean IC_{50} values in the RC rat-only weight regression excluding substances with specific mechanisms of toxicity.

534 [7]Difference between mean animal use with default starting dose and mean animal use with NRU-based LD_{50}. Differences denoted by * were statistically
535 significant (i.e., $p < 0.05$) by a one-sided Wilcoxon signed rank test. Percent difference is shown in parentheses.
536 [8]Proportion of substances for which the GHS acute oral toxicity category (UN 2005) predicted by the *in vitro* NRU test methods matched the *in vivo* category
537 (from **Table 6-6**).
538

For the dose-response slope of 8.3, mean animal savings for the 3T3 NRU ranged from -0.06 (-0.9%) to 2.01 (30.1%) while mean animal savings for the NHK NRU ranged from -0.19 (-2.6%) to 1.88 (28.4%).

For both *in vitro* NRU cytotoxicity test methods, no mean animal savings were observed for substances with $50 < LD_{50} \leq 300$ mg/kg. In fact, slightly more animals were used than when using the default starting dose (i.e., animal savings were negative; -0.02 to -0.16 animal). Since this category includes the default starting dose of 175 mg/kg, animal savings were not expected for this category since savings were determined by comparing animal use with the NRU-based starting dose with animal use at the default starting dose. For the NHK NRU test method, more animals were also used for substances with $LD_{50} \leq 5$ mg/kg (i.e. animal savings were -0.18 to -0.35 animals). The exceptions for having little to no animal savings for the high toxicity substances was for the substances in the $5 < LD_{50} \leq 50$ mg/kg category for the NHK NRU (0.89 [10.9%] to 0.98 [10.6%] animals), but only the 0.98 animals at dose-response = 2 was statistically significant.

Table 10-5 also shows that mean animal savings did not correlate with the accuracy of the GHS acute oral toxicity category predictions (see **Section 6.3**). The toxicity categories with the highest animal savings had low accuracy. Substances with $LD_{50} > 5000$ mg/kg had relatively low accuracy (15 - 25%) for GHS acute oral toxicity category prediction, but the animal savings were relatively high (1.88 - 2.01 animals). For the NHK NRU, substances in the $5 < LD_{50} \leq 50$ mg/kg category had very low accuracy (14%) for GHS acute oral toxicity category prediction, but the animal savings were statistically significant (0.98 animals at dose-response = 2). Possibly the difference between the predicted starting dose and the true LD_{50} vs. the difference between the default starting dose and the true LD_{50} has more influence on animal savings than the accuracy of the LD_{50} prediction. The RC rat-only weight regression excluding substances with specific mechanisms of toxicity improved accuracy (compared with the RC millimole regression) and animal savings for the GHS toxicity categories for substances in the $2000 < LD_{50} \leq 5000$ mg/kg and $LD_{50} > 5000$ mg/kg categories. For substances in the $2000 < LD_{50} \leq 5000$ mg/kg category, accuracy increased

from 0 - 9% (both *in vitro* test methods and dose-response slopes) to 44 - 67% and animal savings increased from 1.25 -1.52 animals to 1.75 - 2.22 animals. For substances with $LD_{50} > 5000$ mg/kg, accuracy increased from 0 - 10% (both *in vitro* NRU test methods and dose-response slopes) to 15 - 25% and animal savings increased from 1.35 - 1.70 animals to 1.77 - 2.01 animals. The RC rat-only weight regression excluding substances with specific mechanisms of toxicity, however, also improved animal savings for substances in the $300 < LD_{50} \leq 2000$ mg/kg toxicity category while which accuracy was decreased compared with the RC millimole regression. Animal savings for substances in the $300 < LD_{50} \leq 2000$ mg/kg toxicity category improved from 0.60 - 0.84 animals (for both *in vitro* NRU test methods and dose-response slopes) to 0.69 - 1.00 animals while accuracy decreased from 100% to 78 - 89%.

10.2.4 Refinement of Animal Use for the UDP When Using 3T3 and NHK NRU-Based Starting Doses

A test method refines animal use when it lessens or eliminates pain or distress in animals or enhances animal well-being (ICCVAM 2003). This section evaluates whether the use of 3T3 and NHK NRU-based starting doses refines animal use by reducing the number of animals that die (i.e., experience pain and distress) during UDP testing compared to the number of animals that die when using the default starting dose of 175 mg/kg. **Table 10-6** reports the refinement results for the UDP simulation modeling using the 5000 mg/kg limit dose. For every regression evaluated, the mean number of deaths when using the NRU-based starting doses was slightly lower than the mean number of deaths when using the default starting dose by approximately 0.1 to 0.2 deaths. The percentage of deaths, however, was slightly higher for the NRU-based starting doses than for the default starting dose since the total number of animals used was lower for the NRU-based starting doses. In general, fewer animals were used and fewer animals died when using an NRU-based starting dose compared with use of the default starting dose.

Table 10-6 Animal Deaths[1] for the UDP[2] Using Starting Doses Based on the 3T3 and NHK NRU Test Methods

Assay/Regression	With Default Starting Dose[3]			With NRU-Based Starting Dose[4]		
	Used	Dead	% Deaths	Used	Dead	% Deaths
3T3 NRU Test Method	**Dose-Response Slope = 2**					
RC millimole[5]	9.77	4.16	42.6%	8.79	3.95	44.9%
RC rat-only weight[6]	9.79	4.18	42.6%	8.66	3.91	45.2%
RC rat-only weight excluding substances with specific mechanisms of toxicity[7]	9.80	4.18	42.7%	8.64	4.03	46.6%
	Dose-Response Slope = 8					
RC millimole[5]	8.11	3.43	42.3%	7.20	3.26	45.3%
RC rat-only weight[6]	8.14	3.44	42.3%	7.11	3.24	45.6%
RC rat-only weight excluding substances with specific mechanisms of toxicity[7]	8.16	3.45	42.3%	7.08	3.34	47.2%
NHK NRU Test Method	**Dose-Response Slope = 2**					
RC millimole [5]	9.75	4.10	42.0%	8.93	3.96	44.3%
RC rat-only weight[6]	9.77	4.11	42.0%	8.83	3.93	44.5%
RC rat-only weight excluding substances with specific mechanisms of toxicity[7]	9.78	4.12	42.1%	8.73	3.99	45.8%
	Dose-Response Slope = 8					
RC millimole [5]	8.12	3.38	41.7%	7.33	3.26	44.5%
RC rat-only weight$_6$	8.14	3.39	41.7%	7.25	3.24	44.7%
RC rat-only weight excluding substances with specific mechanisms of action[7]	8.15	3.40	41.7%	7.15	3.29	46.1%

[1]Numbers are mean numbers of animals used for 2000 simulations for each substance. Although the simulations used whole animals, averaging the results produced fractional numbers of animals. Upper limit dose = 5000 mg/kg. Results are provided for 46 substances in the 3T3 NRU and 47 substances in the NHK NRU test methods.

[2]OECD (2001a); EPA (2002a).

[3]Default starting dose = 175 mg/kg.

[4]Starting dose was one default dose lower than NRU-predicted LD_{50} calculated using the geometric mean of laboratory mean IC_{50} values in the regression specified.

[5]$\log LD_{50}$ (mmol/kg) = 0.435 $\log IC_{50}$ (mM) + 0.625

[6]$\log LD_{50}$ (mg/kg) = 0.372 $\log IC_{50}$ (μg/mL) + 2.024

[7]$\log LD_{50}$ (mg/kg) = 0.357 $\log IC_{50}$ (μg/mL) + 2.194

10.3 Reduction and Refinement of Animal Use for the ATC

10.3.1 Procedure for *In Vivo* Testing Using the ATC

This section describes the general dosing procedure for the conduct of the ATC assay (OECD 2001d). The purpose of the ATC is to classify a test substance into the appropriate GHS category for acute oral toxicity for classification and labeling. This is done by estimating the range of the LD_{50} values for a test substance rather than calculating a point estimate of the LD_{50}. The time between doses is determined by the onset, duration, and severity of toxic signs. Guidance on the type of animals to use, animal housing, clinical observations, etc., which are outside the scope of the current discussion, are provided in the test guidelines (See **Appendix M**).

Main Test

The ATC is based on the stepwise administration of test substances to three animals at a time at one of a number of fixed doses: 5, 50, 300, and 2000 mg/kg (and 5000 mg/kg, if necessary). The starting dose is selected so that at least some of the animals die at that dose. If no information on which to base a starting dose is available, the default starting dose of 300 mg/kg is used. The next step, which may be to stop testing, test at the same dose, test at the next higher dose, or test at the next lower dose, is determined by the starting dose and the outcome of the three animals tested at the starting dose. For example, if the starting dose is 300 mg/kg and two to three animals die or are in a moribund state, the next step is to administer 50 mg/kg to three more animals. However, if zero to one animal dies at 300 mg/kg, three more animals are tested at 300 mg/kg. Most substances required two to four dose steps for substance classification. See **Appendix M** for the outcome-based testing sequence for each starting dose.

Limit Test

For test substances that are likely to be nontoxic, the ATC test method guideline includes a limit test in which six animals (three animals per step) are tested at the limit dose of 2000 mg/kg or 5000 mg/kg (OECD 2001d).

10.3.2 Procedure for Computer Simulation Modeling of the ATC

The simulation process for the ATC was performed using MATLAB® (The MathWorks, Inc. 1996-2004) computational software, which is functionally comparable to SAS® version 8. Two thousand simulations of ATC testing were run for each substance, NRU test method, and dose-mortality slope using an upper limit dose of 2000 mg/kg. The simulation process implements the distributional assumptions underlying the dose-mortality response. The lowest dose at which an animal dies in response to the administration of a toxic substance varies from animal to animal. For an entire population of animals, mortality is assumed to have a log-normal distribution with the mean equal to the log of the true LD_{50}. Sigma (σ), the variability of the simulated population, is the inverse of the slope of the dose-mortality curve. For any given dose, the probability that an animal will die is computed by the cumulative log-normal distribution:

$$\text{Probability (death)} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\log dose} e^{\frac{-(t-\log trueLD_{50})^2}{2\sigma^2}} dt$$

Due to a lack of information for the real dose-mortality curves, the simulations assumed several different values of the slope (i.e., the inverse of σ). Dose-mortality slopes of 0.5, 0.8, 2, 4, and 8.3 were chosen to be comparable to those chosen for simulation modeling of the UDP (see **Section 10.2.2**).

To model the variability of the NRU IC_{50} values within and between laboratories, the values were log-transformed to normalize the distribution of values for each substance. The mean and variance of these log-transformed values were used to generate a log-normal distribution from which to randomly select an IC_{50} value.

The simulation procedure used the following steps for each substance:

1. The LD_{50} value (in mg/kg) from **Table 4-2** was entered as the true LD_{50} value and the choices of assumed slope were entered as the true slope for the dose-mortality curve.

2. An IC_{50} value was selected from a distribution identified by the mean and variance of the IC_{50} values computed from the data to reflect that different laboratories produce different IC_{50} values in different situations (see **Table 5-3** for mean IC_{50} values and standard deviations).
3. The IC_{50} value from Step 2 was used in the regression model being evaluated to compute a predicted LD_{50} value to use as the starting dose.
4. The dosing simulation (of 2000 iterations) was run twice: once with the default starting dose of 300 mg/kg and once with a starting dose equal to the next fixed dose below the LD_{50} estimated by the regression being evaluated (i.e., the NRU-based starting dose). If the NRU-based starting dose was greater than the 2000 mg/kg limit dose, then testing proceeded using the 2000 mg/kg limit test rather than the main test.
5. For every dose group of three animals, one observation was sampled from a binomial distribution with the probability of death calculated by the probability equation for a population of three. The sampled value, referred to as N1, indicates the number of animals, 0, 1, 2, or 3, in the dosing group that die.
6. If $N1 \leq 1$, step 4 is repeated with the same dose. Now the sampled value from the binomial distribution is referred to as N2.
7. If $N2 \leq 1$ and the dose is the highest dose tested, or the dose has already been decreased, the toxicity category is assigned and testing is terminated. If the dose is not the highest dose tested, or if the dose has not been decreased, the dose is increased to the next fixed dose and step 4 is repeated.
8. If $N1 > 1$ or $N2 > 2$, and the dose is the lowest dose tested, or if the dose has already been increased, the toxicity category is assigned and testing is terminated. If the dose is not the lowest dose tested, or if the dose has not already been increased, the dose is decreased to the next fixed dose and step 4 is repeated.

10.3.3 Animal Savings for the ATC When Using 3T3 and NHK NRU-Based Starting Doses

10.3.3.1 *The Effect of Dose-Response Slope on Animal Use*

As described in **Section 10.3.2**, the simulation modeling of animal use for the ATC used five different dose-mortality slopes to assess animal use under various conditions of population variability. **Table 10-7** shows how animal use for the simulated ATC changes with dose-response slope and mean animal use for ATC simulations when using the default starting dose of 300 mg/kg and when using a starting dose that was one fixed dose lower than that predicted by the 3T3 and NHK NRU IC_{50} values with the RC millimole regression. The mean number of animals used for the ATC decreased slightly with increasing slope for both the default starting dose and the NRU-determined starting dose.

Table 10-7 Change in Animal Use[1] with Dose-Response Slope for the ATC[2]

Dose-Response Slope	Default Starting Dose[1,3]	NRU-Based Starting Dose[1,4]	Animals Saved[5]
3T3 NRU Test Method			
0.5	11.10 ± 0.07	10.11 ± 0.24	0.99* (8.9%)
0.8	10.98 ± 0.10	9.95 ± 0.27	1.03* (9.4%)
2.0	10.90 ± 0.16	9.76 ± 0.33	1.13* (10.4%)
4.0	10.84 ± 0.19	9.66 ± 0.35	1.17* (10.8%)
8.3	10.81 ± 0.21	9.64 ± 0.36	1.17* (10.8%)
NHK NRU Test Method			
0.5	11.10 ± 0.07	10.07 ± 0.22	1.03* (9.3%)
0.8	11.00 ± 0.09	9.90 ± 0.24	1.10* (10.0%)
2.0	10.93 ± 0.16	9.72 ± 0.30	1.21* (11.1%)
4.0	10.87 ± 0.19	9.61 ± 0.32	1.26* (11.6%)
8.3	10.84 ± 0.21	9.57 ± 0.34	1.27* (11.7%)

[1]Numbers are mean numbers of animals used and standard errors for 2000 simulations for 46 substances for the 3T3 NRU test method and 47 substances for the NHK NRU test method. Although the simulations used whole animals, averaging the results produced fractional numbers of animals. Limit dose = 2000 mg/kg.
[2]OECD (2001d).
[3]Default starting dose = 300 mg/kg.
[4]Next fixed dose lower than the predicted LD_{50} calculated using the geometric mean of laboratory mean IC_{50} values in the RC millimole regression: $\log LD_{50}$ (mmol/kg) = 0.435 $\log IC_{50}$ (mM) + 0.625.
[5]Difference between mean animal use with default starting dose and mean animal use with NRU-based starting dose. Differences that were statistically significant (i.e., $p < 0.05$) by a one-sided Wilcoxon rank test are noted by *. Percent difference is shown in parentheses.

The mean numbers of animals saved, which was statistically significant (i.e., $p < 0.05$ by one-sided Wilcoxon signed rank tests) when compared with mean animal use for the default

dose, generally increased with increasing slope. To simplify the presentation of animal savings and comparison of the various regressions and starting doses, future results in **Section 10.3.3** will be shown only for dose-response slopes of 2 and 8.3. Results for the other dose-mortality slopes are presented in **Appendices N4-N6.**

10.3.3.2 *Mean Animal Use for ATC Simulations of Testing the NICEATM/ECVAM Reference Substances – Comparison of Regressions and 3T3 and NHK NRU Test Methods*

Table 10-8 shows the mean animal use for testing the NICEATM/ECVAM reference substances using the simulated ATC method when the starting dose was the default starting dose and when the starting dose was one fixed dose lower than that determined by the LD_{50} predicted from the 3T3 and NHK NRU test methods and the regressions (shown in **Table 6-2**) evaluated in **Section 6.3** for prediction of GHS acute oral toxicity category. The mean difference in animal use between the two starting doses is the mean animal savings. All mean differences (i.e., mean animal savings) were statistically significant (i.e., $p < 0.05$ using one-sided Wilcoxon signed rank tests). Mean animal savings ranged from 1.13 (10.4%) to 2.28 (21.1%) animals depending upon the test method, regression, and dose-response slope. The lowest mean animal savings were obtained for the RC millimole regression (1.13 [10.4%] to 1.27 [11.7%] animals) and the highest mean animal savings were obtained with the RC rat-only regression excluding substances with specific mechanisms of toxicity (1.68 [15.4%] to 2.28 [21.1%] animals).

10.3.3.3 *Animal Savings for the ATC by Toxicity Category Using 3T3 and NHK NRU-Based Starting Doses*

Tables 10-9 through **10-11** show mean animal use and mean animal savings for the ATC when used with the *in vitro* NRU cytotoxicity test methods, organized by GHS toxicity category (UN 2005), and when based on the:

- RC millimole regression (**Table 10-9**)
- RC rat-only weight regression (**Table 10-10)**
- RC rat-only weight regression excluding substances with specific mechanisms of toxicity (**Table 10-11)**

753 **Table 10-8 Animal Use[1] for the ATC[2] Using Starting Doses Based on NRU Test Methods with Various Regressions**

Assay/Regression	With Default Starting Dose[3]	With NRU-Based Starting Dose[4]	Animals Saved[5]	With Default Starting Dose[3]	With NRU-Based Starting Dose[5]	Animals Saved[5]	Accuracy[6]
3T3 NRU Test Method	**Dose-Response Slope = 2**			**Dose-Response Slope = 8.3**			
RC millimole[7]	10.90 ± 0.16	9.76 ± 0.33	1.13* (10.4%)	10.81 ± 0.21	9.64 ± 0.36	1.17* (10.8%)	26%
RC rat-only weight[8]	10.90 ± 0.16	9.21 ± 0.31	1.68* (15.5%)	10.81 ± 0.21	8.84 ± 0.36	1.97* (18.2%)	35%
RC rat-only weight excluding substances with specific mechanisms of toxicity[9]	10.90 ± 0.16	9.00 ± 0.29	1.90* (17.4%)	10.81 ± 0.21	8.53 ± 0.33	2.28* (21.1%)	46%
NHK NRU Test Method	**Dose-Response Slope = 2**			**Dose-Response Slope = 8.3**			
RC millimole[7]	10.93 ± 0.16	9.72 ± 0.30	1.21* (11.1%)	10.84 ± 0.21	9.57± 0.34	1.27* (11.7%)	28%
RC rat-only weight[8]	10.93 ± 0.16	9.45 ± 0.30	1.49* (13.6%)	10.84 ± 0.21	9.22 ± 0.34	1.62* (14.9%)	30%
RC rat-only weight excluding substances with specific mechanisms of toxicity[9]	10.93 ± 0.16	9.25 ± 0.26	1.68* (15.4%)	10.84 ± 0.21	8.91 ± 0.31	1.94* (17.9%)	38%

754 [1]Numbers are mean numbers of animals used and standard errors for 2000 ATC simulations each for 46 substances for the 3T3 NRU test method and 47
755 substances for the NHK NRU test method. Limit dose = 2000 mg/kg
756 [2]OECD (2001d).
757 [3]Default starting dose = 300 mg/kg.
758 [4]Starting dose was one fixed dose lower than NRU-predicted LD_{50} calculated using the geometric mean of laboratory mean IC_{50} values in the regression
759 specified.
760 [5]Difference between mean animal use with default starting dose and mean animal use with NRU-based LD_{50}. Percentage difference is shown in parentheses.
761 Differences marked by * were statistically significant (i.e., $p < 0.05$) using a one-sided Wilcoxon signed rank test.
762 [6]Proportion of substances for which the GHS acute oral toxicity category (UN 2005) predicted by the *in vitro* NRU test methods matched the *in vivo* category
763 (from **Tables 6-4** to **6-6**).
764 [7]$\log LD_{50}$ (mmol/kg) = 0.435 $\log IC_{50}$ (mM) + 0.625.
765 [8]$\log LD_{50}$ (mg/kg) = 0.372 $\log IC_{50}$ (μg/mL) + 2.024.
766 [9]$\log LD_{50}$ (mg/kg) = 0.357 $\log IC_{50}$ (μg/mL) + 2.194.

The summarized data come from the same analyses as the data provided in **Table 10-8**.

Consistencies noted in the mean animal savings data provided in the tables included:

- For each test method and regression, the highest mean animal savings were generally in the $LD_{50} \leq 5$ mg/kg and $LD_{50} > 5000$ mg/kg toxicity categories.
- For each test method and regression, the lowest mean animal savings were in the $50 < LD_{50} \leq 300$ mg/kg and $300 < LD_{50} \leq 2000$ mg/kg toxicity categories.

Animal Savings for the ATC by Toxicity Category Using 3T3 and NHK NRU-Based Starting Doses with the RC Millimole Regression

Table 10-9 shows the mean animal savings for the ATC by GHS toxicity category for the *in vitro* NRU test methods used with the RC millimole regression. Mean differences between animal use for the default starting dose and animal use with the NRU-determined starting dose were statistically significant (i.e., $p < 0.05$) by a one-sided Wilcoxon signed rank test for the following GHS toxicity categories, test methods, and dose-response slopes:

- $LD_{50} \leq 5$ mg/kg for the 3T3 NRU at both dose-response slopes (2.75 [29.5%] to 2.80 [31.1%] animals)
- $2000 < LD_{50} \leq 5000$ mg/kg for the 3T3 NRU at dose-response slope = 8 (0.35 [2.9%] animals) and for the NHK NRU at dose-response slope = 2 (0.38 [3.4%] animals)
- $LD_{50} > 5000$ mg/kg for the 3T3 NRU at both dose-response slopes (2.32 [29.6%] and 2.46 [20.5%] animals) and for the NHK NRU at dose-response slope = 2 (2.34 [19.7%] animals)

For the dose-response slope of 2, mean animal savings for the 3T3 NRU test method ranged from -0.24 (-2.5%) to 2.75 (29.5%) animals while animal savings for the NHK NRU test method ranged from -0.02 (-0.2%) to 2.43 (19.9%) animals. For the dose-response slope of 8.3, mean animal savings for the 3T3 NRU test method ranged from -0.47 (-5.1%) to 2.80 (31.1%) animals while mean animal savings for the NHK NRU test method ranged from -0.23 (-2.4%) to 2.79 (23.0%) animals.

Table 10-9 Animal Savings[1] for the ATC[2] by GHS Toxicity Category[3] Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Millimole Regression[4]

		Dose-Response Slope = 2			Dose-Response Slope = 8.3			
Toxicity Category[3]	**Number of Reference Substances**	**With Default Starting Dose[5]**	**With NRU-Based Starting Dose[6]**	**Animals Saved[7]**	**With Default Starting Dose[5]**	**With NRU-Based Starting Dose[6]**	**Animals Saved[7]**	**Accuracy[8]**
		3T3 NRU Test Method						
$LD_{50} \leq 5$ mg/kg	7	9.35 ± 0.11	6.60 ± 0.87	2.75* (29.5%)	9.00 ± 0.001	6.20 ± 0.88	2.80* (31.1%)	0%
$5 < LD_{50} \leq 50$ mg/kg	6	12.22 ± 0.05	11.12 ± 0.94	1.10 (9.0%)	12.13 ± 0.08	10.71 ± 1.00	1.42 (11.7%)	17%
$50 < LD_{50} \leq 300$ mg/kg	6	10.70 ± 0.37	10.01 ± 0.08	0.69 (6.5%)	9.72 ± 0.48	9.39 ± 0.16	0.32 (3.3%)	67%
$300 < LD_{50} \leq 2000$ mg/kg	6	9.79 ± 0.08	10.04 ± 0.14	-0.24 (-2.5%)	9.20 ± 0.11	9.67 ± 0.27	-0.47 (-5.1%)	100%
$2000 < LD_{50} \leq 5000$ mg/kg	11	11.18 ± 0.08	11.02 ± 0.13	0.16 (1.4%)	11.90 ± 0.04	11.55 ± 0.20	0.35* (2.9%)	0%
$LD_{50} > 5000$ mg/kg	10	11.90 ± 0.03	9.58 ± 0.91	2.32* (19.5%)	12.00 ± 0.000	9.54 ± 0.97	2.46* (20.5%)	10%
		NHK NRU Test Method						
$LD_{50} \leq 5$ mg/kg	7	9.37 ± 0.12	7.62 ± 1.12	1.76 (18.7%)	9.00 ± 0.002	7.25 ± 1.04	1.75 (19.5%)	0%
$5 < LD_{50} \leq 50$ mg/kg	6	12.2 ± 0.04	9.77 ± 0.34	2.43 (19.9%)	12.14 ± 0.09	9.35 ± 0.18	2.79 (23.0%)	50%
$50 < LD_{50} \leq 300$ mg/kg	6	10.75 ± 0.39	10.32 ± 0.36	0.43 (4.0%)	9.74 ± 0.49	9.97 ± 0.78	-0.23 (-2.4%)	50%
$300 < LD_{50} \leq 2000$ mg/kg	6	9.79 ± 0.08	9.81 ± 0.08	-0.02 (-0.2%)	9.21 ± 0.13	9.28 ± 0.13	-0.06 (-0.7%)	100%
$2000 < LD_{50} \leq 5000$ mg/kg	11	11.19 ± 0.09	10.81 ± 0.27	0.38* (3.4%)	11.90 ± 0.04	11.17 ± 0.73	0.73 (6.2%)	9%
$LD_{50} > 5000$ mg/kg	11	11.92 ± 0.02	9.58 ± 0.85	2.34* (19.7%)	12.00 ± 0.000	9.52 ± 0.90	2.48 (20.6%)	0%

[1]Numbers are mean number of animals used and standard errors for 2000 simulations for each substance with a limit dose of 2000 mg/kg. Results are provided for 46 substances in the 3T3 NRU test method and 47 substances in the NHK NRU test method categorized using the initial LD_{50} values from **Table 3-2**. Although the simulations used whole animals, averaging the results produced fractional numbers of animals.

[2]OECD (2001d).

[3]GHS-Globally Harmonized System of Classification and Labelling of Chemicals with LD_{50} in mg/kg (UN 2005).

[4]RC millimole regression is log LD_{50} (mmol/kg) = 0.435 log IC_{50} (mM) + 0.625.

[5]Default starting dose = 300 mg/kg.

[6]Starting dose was the next fixed dose lower than the predicted LD_{50} from using the NRU IC_{50} in the RC millimole regression.

[7]Difference between mean animal use with default starting dose and mean animal use with NRU-based starting dose. Statistically significant differences (i.e., $p < 0.05$) by a one-sided Wilcoxon signed rank test are noted by *. Percentage difference is shown in parentheses.

[8]Proportion of substances for which the GHS acute oral toxicity category (UN 2005) predicted by the *in vitro* NRU test methods matched the *in vivo* category (from **Table 6-4**).

.

For both dose-response slopes, the mean animal savings using the 3T3 NRU test method was lower than the mean animal savings using the NHK NRU test method for substances in four of the six toxicity categories: $5 < LD_{50} \leq 50$ mg/kg; $3000 < LD_{50} \leq 2000$; $2000 < LD_{50} \leq 5000$ mg/kg; and $LD_{50} > 5000$ mg/kg. Mean animal savings using the 3T3 NRU test method was higher than the mean animal savings using the NHK NRU test method for substances in the other two toxicity categories: $LD_{50} \leq 5$ mg/kg and $50 < LD_{50} \leq 300$ mg/kg. For the 3T3 NRU test method, the highest mean animal savings occurred for substances in the category for $LD_{50} \leq 5$ mg/kg (23.2 [19.5%] animals at dose-response slope = 2 and 2.46 [20.5%] animals at dose-response slope = 8.3). For the NHK NRU test method, the highest mean animal savings occurred for substances in the category for $5 < LD_{50} \leq 50$ mg/kg (2.43 [19.9%] animals at dose-response slope = 2 and 2.79 [23.0%] animals at dose-response slope = 8.3); however, the animal savings were not statistically significant.

For both test methods, the smallest mean animal savings (≤ 0.69) were observed for substances with LD_{50} values between 50 and 2000 mg/kg. Since the default starting dose was 300 mg/kg, little change in mean animal use was expected for substances in the $50 < LD_{50} \leq 300$ mg/kg and $300 < LD_{50} \leq 2000$ mg/kg categories. For both test methods and dose-response slopes, mean animal savings for the substances in the $50 < LD_{50} \leq 300$ mg/kg category were -0.23 to 0.69 animals. For both test methods and dose-response slopes, there were no animal savings for substances in the $300 < LD_{50} \leq 2000$ mg/kg category. In fact, slight more animals were used for the NRU-based starting doses than for the default starting dose (-0.02 to -0.47 animals).

Table 10-9 also shows that mean animal savings did not correlate with the accuracy of the GHS acute oral toxicity category predictions (see **Section 6.3**). The toxicity categories with the highest animal savings had low accuracy. The 3T3 NRU test method produced the highest animal savings (2.75 - 2.80) for substances with $LD_{50} \leq 5$ mg/kg, which had 0% accuracy for GHS acute oral toxicity category prediction. Substances in the $300 < LD_{50} \leq 2000$ mg/kg category had 100% accuracy for GHS acute oral toxicity category prediction, but had no animal savings (≤ 0.2 animals). Possibly the difference between the

predicted starting dose and the true LD_{50} vs. the difference between the default starting dose and the true LD_{50} has more influence on animal savings than the accuracy of the LD_{50} prediction.

Animal Savings for the ATC by Toxicity Category Using 3T3 and NHK NRU-Based Starting Doses with the RC Rat-Only Weight Regression

Table 10-10 shows the animal savings for the simulation ATC method by GHS toxicity category for the *in vitro* NRU cytotoxicity test methods used with the RC rat-only weight regression. Mean animal savings were statistically significant (i.e., $p < 0.05$) by a one-tailed Wilcoxon signed rank test for the following GHS toxicity categories, test methods, and dose-response slopes:

- $LD_{50} \leq 5$ mg/kg for both NRU test methods and dose-response slopes (2.03 [21.9%] to 2.57 [28.5%] animals)
- $2000 < LD_{50} \leq 5000$ mg/kg for the 3T3 NRU test method at both dose-response slopes (1.39 [12.4%] to 2.56 [21.5%] animals)
- $LD_{50} > 5000$ mg/kg for both NRU test methods and dose-response slopes (2.92 [24.5%] to 3.5 [29.2%] animals)

For the 3T3 NRU and NHK NRU test methods, mean animal savings were similar for most toxicity categories at both dose-response slopes, with the mean savings for the 3T3 NRU slightly higher than that for the NHK NRU for most toxicity categories. For the dose-response slope of 2, mean animal savings for the 3T3 NRU test method (for the various toxicity categories) ranged from -0.32 (-3.3%) to 32.8 (27.5%) animals while mean animal savings for the NHK NRU test method ranged from 0.03 (0.3%) to 2.92 (24.5%) animals. For the dose-response slope of 8.3, animal savings for the 3T3 NRU test method ranged from -0.63 (-6.8%) to 3.50 (29.2%) animals while mean animal savings for the NHK NRU test method ranged from -0.23 (-2.4%) to 3.12 (26.0%) animals.

For both test methods, there were no mean animal savings (≤ 0.03 animals) for substances with LD_{50} values between 300 and 2000 mg/kg. For both test methods and dose-response slopes, mean animal savings for the substances in the $50 < LD_{50} \leq 300$ mg/kg category were

Table 10-10 Animal Savings[1] for the ATC[2] by GHS Toxicity Category[3] Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Rat-Only Weight Regression[4]

		Dose-Response Slope = 2			Dose-Response Slope = 8.3			
Toxicity Category[3]	**Number of Reference Substances**	**With Default Starting Dose[5]**	**With NRU-Based Starting Dose[6]**	**Animals Saved[7]**	**With Default Starting Dose[5]**	**With NRU-Based Starting Dose[6]**	**Animals Saved[7]**	**Accuracy[8]**
		3T3 NRU Test Method						
$LD_{50} \leq 5$ mg/kg	4	9.35 ± 0.11	6.83 ± 0.84	2.52* (27.0%)	9.00 ± 0.001	6.43 ± 0.85	2.57* (28.5%)	0%
$> 5 < LD_{50} \leq 50$ mg/kg	7	12.22 ± 0.05	10.33 ± 0.52	1.88 (15.4%)	12.13 ± 0.08	9.94 ± 0.54	2.20 (18.1%)	14%
$> 50 < LD_{50} \leq 300$ mg/kg	5	10.70 ± 0.37	9.94 ± 0.10	0.76 (7.1%)	9.72 ± 0.48	9.23 ± 0.12	0.48 (5.0%)	80%
$> 300 < LD_{50} \leq 2000$ mg/kg	9	9.79 ± 0.08	10.11 ± 0.29	-0.32 (-3.3%)	9.20 ± 0.11	9.83 ± 0.55	-0.63 (-6.8%)	78%
$> 2000 < LD_{50} \leq 5000$ mg/kg	9	11.18 ± 0.08	9.79 ± 0.47	1.39* (12.4%)	11.9 ± 0.04	9.34 ± 0.82	2.56* (21.5%)	44%
> 5000 mg/kg	12	11.90 ± 0.03	8.62 ± 0.94	3.28* (27.5%)	12.00 ± 0.00	8.50 ± 0.99	3.50* (29.2%)	0%
		NHK NRU Test Method						
$LD_{50} \leq 5$ mg/kg	4	9.37 ± 0.12	7.32 ± 0.88	2.05* (21.9%)	9.00 ± 0.002	6.97 ± 0.81	2.03* (22.6%)	0%
$> 5 < LD_{50} \leq 50$ mg/kg	7	12.20 ± 0.04	9.72 ± 0.30	2.48 (20.3%)	12.14 ± 0.08	9.35 ± 0.17	2.79 (23.0%)	14%
$> 50 < LD_{50} \leq 300$ mg/kg	5	10.75 ± 0.39	10.30 ± 0.34	0.45 (4.2%)	9.74 ± 0.49	9.97 ± 0.78	-0.23 (-2.4%)	60%
$> 300 < LD_{50} \leq 2000$ mg/kg	9	9.79 ± 0.08	9.76 ± 0.08	0.03 (0.3%)	9.21 ± 0.13	9.20 ± 0.11	0.02 (0.2%)	89%
$> 2000 < LD_{50} \leq 5000$ mg/kg	9	11.19 ± 0.09	10.45 ± 0.40	0.73 (6.6%)	11.90 ± 0.04	10.55 ± 0.69	1.35 (11.3%)	11%
$LD_{50} > 5000$ mg/kg	13	11.92 ± 0.02	9.00 ± 0.88	2.92* (24.5%)	12.00 ± 0.00	8.88 ± 0.93	3.12* (26.0%)	8%

[1]Numbers are mean number of animals used and standard errors for 2000 simulations for each substance with a limit dose of 5000 mg/kg. Although the simulations used whole animals, averaging the results produced fractional numbers of animals. Results are provided for 46 substances in the 3T3 NRU test method and 47 substances in the NHK NRU test method categorized using the reference LD_{50} values from **Table 4-2**.

[2]OECD (2001d).

[3]GHS-Globally Harmonized System of Classification and Labelling of Chemicals with LD_{50} in mg/kg (UN 2005).

[4]From **Table 6-2;** $\log LD_{50}$ (mg/kg) = 0.372 $\log IC_{50}$ (μg/mL) + 2.024

[5]Default starting dose = 300 mg/kg.

[6]Starting dose was one fixed dose lower than the NRU-predicted LD_{50} calculated using the NRU IC_{50} in the RC rat-only weight regression.

[7]Difference between mean animal use with default starting dose and mean animal use with NRU-based LD_{50}. Differences marked by * were statistically significant (i.e., $p < 0.05$) by a one-sided Wilcoxon signed rank test. Percentage difference is shown in parentheses.

[8]Proportion of substances for which the GHS acute oral toxicity category (UN 2005) predicted by the *in vitro* NRU test methods matched the *in vivo* category (from **Table 6-5**).

also relatively small (-0.23 to 0.76) animals. Since the default starting dose was 300 mg/kg, little change in mean animal use was expected for substances in the $50 < LD_{50} \leq 300$ mg/kg and $300 < LD_{50} \leq 2000$ mg/kg categories.

Table 10-10 also shows that mean animal savings did not correlate with the accuracy of the GHS acute oral toxicity category predictions (see **Section 6.3**). The toxicity categories with the highest animal savings had low accuracy. For example, animal savings for substances in the $LD_{50} > 5000$ mg/kg category were 2.92 - 3.50 animals (for both *in vitro* NRU test methods and dose-response slopes) and accuracy was 0 - 8%. In addition, substances in toxicity categories with the lowest animal savings had the highest accuracy for GHS acute oral toxicity category prediction. Substances in the $300 < LD_{50} \leq 2000$ mg/kg category had relatively high accuracy for GHS acute oral toxicity category prediction (i.e., 78% for the 3T3 NRU and 89% for the NHK NRU), but had the lowest animal savings (≤ 0.45 animals). Possibly the difference between the predicted starting dose and the true LD_{50} vs. the difference between the default starting dose and the true LD_{50} has more influence on animal savings than the accuracy of the LD_{50} prediction.

Animal Savings for the ATC by Toxicity Category Using 3T3 and NHK NRU-Based Starting Doses with the RC Rat-Only Weight Regression Excluding Substances with Specific Mechanisms of Toxicity

Table 10-11 shows the animal savings by GHS toxicity category for simulated ATC testing using the *in vitro* NRU cytotoxicity test methods with the RC rat-only weight regression excluding substances with specific mechanisms of toxicity. Mean animal savings were statistically significant (i.e., $p < 0.05$) by a one-tailed Wilcoxon signed rank test for the following GHS toxicity categories, test methods, and dose-response slopes:

- $LD_{50} \leq 5$ mg/kg for the 3T3 NRU test method at dose-response slope = 8.3 (2.16 [24.0%] animals) and for the NHK NRU test method at dose-response slope = 2 (1.27 [13.5%] animals)
- $2000 < LD_{50} \leq 5000$ mg/kg for both NRU test methods and both dose-response slopes (1.23 [11.0%] to 3.07 [25.8%] animals)

- LD_{50} > 5000 mg/kg for both NRU test methods and both dose-response slopes (3.79 [31.8%] to 4.04 [33.7%] animals)

For the 3T3 NRU and NHK NRU test methods, mean animal savings were similar for most toxicity categories at both dose-response slopes, with the mean savings for the 3T3 NRU slightly higher than that for the NHK NRU. For the dose-response slope of 2, mean animal savings for the 3T3 NRU test method (for the various toxicity categories) ranged from 0.02 (0.2%) to 4.08 (34.3%) animals while mean animal savings for the NHK NRU test method ranged from 0.00 (0.0%) to 3.79 (31.8%) animals. For the dose-response slope of 8.3, animal savings for the 3T3 NRU test method ranged from -0.03 (-0.4%) to 4.38 (36.5%) animals while mean animal savings for the NHK NRU test method ranged from -0.06 (-0.6%) to 4.04 (33.7%) animals.

For both test methods, there were no mean animal savings (≤ 0.02 animals) for substances with LD_{50} values between 300 and 2000 mg/kg. For both test methods and dose-response slopes, mean animal savings for the substances in the $50 < LD_{50} \leq 300$ mg/kg category were also relatively small (-0.06 to 0.79) animals. Since the default starting dose was 300 mg/kg, little change in mean animal use was expected for substances in the $50 < LD_{50} \leq 300$ mg/kg and $300 < LD_{50} \leq 2000$ mg/kg categories.

Table 10-11 also shows that mean animal savings did not correlate with the accuracy of the GHS acute oral toxicity category predictions (see **Section 6.3**). The toxicity category with the highest animal savings (LD_{50} > 5000 mg/kg) had low accuracy (15 - 25%). Substances in the $300 < LD_{50} \leq 2000$ mg/kg category had very high accuracy, 78-89%, but no animal savings. Perhaps the difference between the predicted starting dose and the true LD_{50} vs. the difference between the default starting dose and the true LD_{50} has more influence on animal savings than the accuracy of the LD_{50} prediction.

Table 10-11 Animal Savings[1] for the ATC[2] By GHS Toxicity Category[3] Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Rat-Only Weight Regression Excluding Substances with Specific Mechanisms of Toxicity[4]

Toxicity Category[3]	Number of Reference Substances	Dose-Response Slope = 2			Dose-Response Slope = 8.3			Accuracy[8]
		With Default Starting Dose[5]	With NRU-Based Starting Dose[6]	Animals Saved[7]	With Default Starting Dose[5]	With NRU-Based Starting Dose[6]	Animals Saved[7]	
		3T3 NRU Test Method						
$LD_{50} \leq 5$ mg/kg	4	9.35 ± 0.11	7.23 ± 0.83	2.12 (22.6%)	9.00 ± 0.001	6.84 ± 0.86	2.16* (24.0%)	0%
$> 5 < LD_{50} \leq 50$ mg/kg	7	12.22 ± 0.05	10.52 ± 0.50	1.70 (13.9%)	12.13 ± 0.08	10.18 ± 0.54	1.96 (16.1%)	14%
$> 50 < LD_{50} \leq 300$ mg/kg	5	10.70 ± 0.37	9.92 ± 0.09	0.79 (7.3%)	9.72 ± 0.48	9.24 ± 0.13	0.48 (4.9%)	80%
$> 300 < LD_{50} \leq 2000$ mg/kg	9	9.79 ± 0.08	9.77 ± 0.07	0.02 (0.2%)	9.20 ± 0.11	9.24 ± 0.13	-0.03 (-0.4%)	78%
$> 2000 < LD_{50} \leq 5000$ mg/kg	9	11.18 ± 0.08	9.50 ± 0.47	1.67* (15.0%)	11.90 ± 0.04	8.83 ± 0.82	3.07* (25.8%)	67%
> 5000 mg/kg	12	11.90 ± 0.03	7.82 ± 0.77	4.08* (34.3%)	12.00 ± 0.00	7.62 ± 0.82	4.38* (36.5%)	25%
		NHK NRU Test Method						
$LD_{50} \leq 5$ mg/kg	4	9.37 ± 0.12	8.11 ± 0.65	1.27* (13.5%)	9.00 ± 0.002	7.76 ± 0.58	1.24 (13.8%)	0%
$> 5 < LD_{50} \leq 50$ mg/kg	7	12.20 ± 0.04	9.87 ± 0.33	2.33 (19.1%)	12.14 ± 0.09	9.52 ± 0.27	2.62 (21.6%)	14%
$> 50 < LD_{50} \leq 300$ mg/kg	5	10.75 ± 0.39	10.19 ± 0.26	0.55 (5.2%)	9.74 ± 0.49	9.80 ± 0.61	-0.06 (-0.6%)	60%
$> 300 < LD_{50} \leq 2000$ mg/kg	9	9.79 ± 0.08	9.79 ± 0.08	0.00 (0.0%)	9.21 ± 0.13	9.21 ± 0.12	0.01 (0.1%)	89%
$> 2000 < LD_{50} \leq 5000$ mg/kg	9	11.19 ± 0.09	9.96 ± 0.45	1.23* (11.0%)	11.90 ± 0.04	9.62 ± 0.80	2.28* (19.2%)	44%
$LD_{50} > 5000$ mg/kg	13	11.92 ± 0.02	8.13 ± 0.76	3.79* (31.8%)	12.00 ± 0.000	7.96 ± 0.81	4.04* (33.7%)	15%

[1]Numbers are mean number of animals used and standard errors for 2000 simulations for each substance with a limit dose of 2000 mg/kg. Although the simulations used whole animals, averaging the results produced fractional numbers of animals. Results are provided for 46 substances in the 3T3 NRU test method and 47 substances in the NHK NRU test method categorized using the reference LD_{50} values from **Table 4-2**.

[2]OECD (2001d).

[3]GHS-Globally Harmonized System of Classification and Labelling of Chemicals with LD_{50} in mg/kg (UN 2005).

[4]From **Table 6-2**; log LD_{50} (mg/kg) = 0.357 log IC_{50} (μg/mL) + 2.194.

[5]Default starting dose = 300 mg/kg.

[6]Starting dose was one fixed dose lower than the NRU-predicted LD_{50} calculated using the NRU IC_{50} in the RC rat-only weight regression excluding substances with specific mechanisms of toxicity.

[7]Difference between mean animal use with default starting dose and mean animal use with NRU-based LD_{50}. Statistically significant differences (i.e., $p < 0.05$) by a one-sided Wilcoxon signed rank test are noted by *. Percentage difference is shown in parentheses.

960 [8]Proportion of substances for which the GHS acute oral toxicity category (UN 2005) predicted by the *in vitro* NRU test methods matched the *in vivo* category
961 (from **Table 6-5**).

The RC rat-only weight regression excluding substances with specific mechanisms of toxicity improved accuracy (compared with the RC millimole regression) and animal savings for the GHS toxicity categories for substances in the $2000 < LD_{50} \leq 5000$ mg/kg and $LD_{50} > 5000$ mg/kg categories. For the $2000 < LD_{50} \leq 5000$ mg/kg category, accuracy improved from 0 - 9% (both *in vitro* NRU test methods) to 44 - 67% and animal savings improved from 0.16 - 0.73 animals to 1.23 - 3.07 animals. For substances with $LD_{50} > 5000$ mg/kg, accuracy improved from 0 - 10% (both *in vitro* NRU test methods) to 15 - 25% and animal savings improved from 2.32 - 2.48 animals to 3.79 - 4.38 animals. Although the RC rat-only weight regression excluding substances with specific mechanisms of toxicity had no animal savings for substances in the $300 < LD_{50} \leq 2000$ mg/kg toxicity category (≤ 0.02 animals), it produced a small improvement over the RC millimole regression since as high as 0.47 more animals were used (compared with the default starting dose).

10.3.4 Refinement of Animal Use for the ATC when using 3T3 and NHK NRU-Based Starting Doses

A test method refines animal use when it lessens or eliminates pain or distress in animals or enhances animal well-being (ICCVAM 2003). This section evaluates whether the use of 3T3 and NHK NRU-based starting doses refines animal use by reducing the number of animals that die during ATC testing compared to the number of animals that die when using the default starting dose of 300 mg/kg. **Table 10-12** reports the refinement results for the ATC simulation modeling using the 2000 mg/kg limit dose. For every regression evaluated, the mean number of deaths when using the 3T3 and NHK NRU-based starting doses was less than the mean number of deaths when using the default starting dose by approximately 0.6 to 0.7 deaths. For the RC millimole regression and the RC rat-only weight regression, the percentage of deaths (compared with the number of animals used) was also slightly lower for the NRU-based starting dose compared with the default starting dose. For the RC rat-only weight regression excluding substances with specific mechanisms of action, the percentage of deaths (compared to the total number of animals used) when using the 3T3 and NHK NRU-based starting doses was about the same as the percentage of deaths when using the default starting dose. In general, fewer animals were used with the NRU-based starting dose and fewer animals died.

Table 10-12 Animal Deaths[1] for the ATC[2] Using Starting Doses Based on the 3T3 and NHK NRU Test Methods

Assay/ Regression	Default Starting Dose[3]			NRU-Based Starting Dose[4]		
	Used	Dead	% Deaths	Used	Dead	% Deaths
3T3 NRU	**Dose-Response Slope = 2**					
RC millimole[5]	10.90	3.55	32.6%	9.76	2.87	29.4%
RC rat-only[6]	10.90	3.55	32.6%	9.21	2.82	30.6%
RC rat-only excluding substances with specific mechanisms of toxicity[7]	10.90	3.55	32.6%	9.00	2.92	32.4%
	Dose-Response Slope = 8.3					
RC millimole [5]	10.81	3.03	28.0%	9.64	2.38	24.7%
RC rat-only[6]	10.81	3.03	28.0%	8.84	2.33	26.3%
RC rat-only excluding substances with specific mechanisms of toxicity[7]	10.81	3.03	28.0%	8.53	2.42	28.3%
NHK NRU	**Dose-Response Slope = 2**					
RC millimole [5]	10.93	3.47	31.8%	9.72	2.82	29.0%
RC rat-only[6]	10.93	3.47	31.8%	9.45	2.78	29.4%
RC rat-only excluding substances specific mechanisms of toxicity[7]	10.93	3.47	31.8%	9.25	2.91	31.5%
	Dose-Response Slope = 8.3					
RC millimole [5]	10.84	2.97	27.4%	9.57	2.34	24.4%
RC rat-only[6]	10.84	2.97	27.4%	9.22	2.30	24.9%
RC rat-only excluding substances with specific mechanisms of toxicity[7]	10.84	2.97	27.4%	8.91	2.43	27.3%

[1]Numbers are mean numbers of animals used for 2000 simulations for each substance (46 substances in the 3T3 NRU test method and 47 substances in the NHK NRU test method). Although the simulations used whole animals, averaging the results produced fractional numbers of animals. Upper limit dose = 2000 mg/kg.
[2]OECD (2001d).
[3]Default starting dose = 300 mg/kg.
[4]Starting dose was one fixed dose lower than the NRU-predicted LD_{50}.
[5]log LD_{50} (mmol/kg) = 0.435 log IC_{50} (mM) + 0.625.
[6]log LD_{50} (mg/kg) = 0.372 log IC_{50} (μg/mL) + 2.024.
[7]log LD_{50} (mmol/kg) = 0.357 log IC_{50} (mM) + 2.194.

10.4 Summary

Computer simulation modeling of UDP testing using the default dose progression shows that, for the subset of NICEATM/ECVAM reference substances evaluated, the prediction of starting doses using the 3T3 and NHK NRU test methods with the RC millimole regression

resulted in the use of statistically ($p < 0.05$) fewer animals for UDP testing by an average of 0.79 - 0.97 animals (8.4 - 11.2%) depending upon the *in vitro* NRU cytotoxicity test method and the dose-response slope (of 2 or 8.3) used. Mean animal savings improved to 1.00 to 1.16 animals (10.7 - 13.3%) for the RC rat-only weight regression excluding substances with specific mechanisms of toxicity.

When reference substances were grouped by GHS toxicity category, there were no mean animal savings for simulated UDP testing for substances with $50 < LD_{50} \leq 300$ mg/kg. Statistically significant animal savings were for observed for substances with $2000 < LD_{50} \leq 5000$ mg/kg and $LD_{50} > 5000$ mg/kg for both NRU test methods. When using the RC millimole regression, animal savings for these categories ranged from 1.25 to 1.70 animals (13.5 to 25.4%). Use of the RC rat-only weight regression excluding substances with specific mechanisms of toxicity improved animal savings for substances in these toxicity categories to 1.75 to 2.22 animals (18.3 to 30.1%). Using the 3T3 and NHK NRU IC_{50} values to estimate starting doses for the simulated UDP also resulted approximately 0.1 to 0.2 fewer mean deaths compared with the use of the default starting dose.

Computer simulation modeling of ATC testing with GHS cut points shows that, for the reference substances tested in this validation study, the prediction of starting doses using the 3T3 and NHK NRU test methods with the RC millimole regression resulted in the use of statistically ($p < 0.05$) fewer animals for ATC testing by an average of 1.13 to 1.27 animals (10.4 - 11.7%) depending upon the *in vitro* NRU cytotoxicity test method and the dose-response slope (of 2 or 8.3) used. Animal savings improved to a mean of 1.68 to 2.28 animals (15.4 - 21.1%) for the RC rat-only weight regression excluding substances with specific mechanisms of toxicity.

When test substances were grouped by GHS toxicity category, mean animal savings for ATC testing using the RC millimole regression were statistically significant for the 3T3 NRU at both dose-response slopes for substances with $LD_{50} \leq 5$ mg/kg (2.75 - 2.80 animals [29.5 - 31.1%]) and for substances with $LD_{50} > 5000$ mg/kg (2.32 [19.5%] - 2.46 [20.5%] animals). Mean ATC animal savings with the RC millimole regression were statistically significant

with the NHK NRU at dose-response = 2 for substances with $2000 < LD_{50} \leq 5000$ mg/kg (0.38 [3.4%] animals) and for substances with $LD_{50} > 5000$ mg/kg (2.34 animals [19.7%]). Using the RC rat-only weight regression excluding substances with specific mechanisms of toxicity, statistically significant animal savings were observed for both test methods and dose response slopes for substances with $2000 < LD_{50} \leq 5000$ mg/kg (1.23 [11.0%] - 3.07 [25.8%] animals) and substances with $LD_{50} > 5000$ mg/kg (3.79 [31.8%] - 4.38 [36.5%] animals). Animal savings were also statistically significant for substances with $LD_{50} \leq 5$ mg/kg using the 3T3 NRU at dose-response slope = 8.3 (2.16 [24.0%]) and using the NHK NRU at dose-response slope = 2 (1.27 [13.5%)]. Using the NRU IC_{50} values to estimate starting doses for the ATC refined animal use by producing approximately 0.6 to 0.7 fewer mean animal deaths than when the default starting dose of 300 mg/kg was used.

Spielmann et al. (1999) indicated that 76% (845/1115) of the industrial substances submitted to the Federal Institute for Health Protection of Consumers and Veterinary Medicine in Berlin, Germany, since 1982 had $LD_{50} > 2000$ mg/kg. Thus, the selection of starting doses using the *in vitro* NRU methods may save a considerable number of animals since animal savings are highest for the least toxic substances. However, the extent to which these substances represent the world of substances in commerce is not known.

[This Page Intentionally Left Blank]